Understanding International Price Dispersion through Export Data from China

Ao Dai

Honors Thesis

Advised by Dr. Joel Rodrigue

Spring 2017

I. Introduction and Related Literature

The Law-of-One-Price (LOP) states that identical goods in different countries should have identical prices if the prices are expressed in common currency units. Purchasing Power Parity (PPP), which states that similar baskets of goods should cost the same once expressed in common units, is the theory that supports this claim. In reality, the prices of the same good traded in different countries and regions usually vary dramatically. In this research, we will be examining several factors that might explain price dispersion in goods traded internationally.

Relatively few empirical studies have tested either The Law-of-One-Price or Purchasing Power Parity using firm-level microdata. Instead, most previous studies have only tested their implication: changes in international relative prices should equal zero. Our research uses the trade data collected by Chinese Customs, which provides export prices of products for different firms manufactured in China and sold to different countries. There are many advantages based on the data source selected by this research. First of all, all prices are measured in USD. Most previous research collected price data denominated in different currencies, and our research has the advantage of using first-hand price data without having to consider the fluctuating nominal exchange rate. Secondly, our data source provides yearly price data for products over a 7-year period from 2000 to 2006, whereas many previous studies compared prices for the same product for different countries with a much larger time interval such as five years. In addition, our data source covers as many as 213 unique countries and 5,144 different products, making the data set very comprehensive as far as number of countries and products are concerned. Lastly, even though the data captures a spatial dimension in the export destinations, it has the advantage of measuring prices in the same country. Since retailers in the export destination usually consider many other cost factors before pricing these products, the data serves as a better reflection of the raw price of the products.

We begin our research by examining the distribution of deviations from The Law-of-One-Price on a good-by-good basis for the 10 most common trade partners of China. Using the average price of the same product across all countries as the numeraire (weighted based on the quantity traded at different prices), we plot the distribution of the deviations for the ten most common-trade partners. The results show that although the range of the deviations for each country can be very large, they tend to be zero on average. In other words, there are as many overpriced goods as underpriced goods in most countries.

The next step we take is to try to explain price in dispersions in two dimensions that we think might be related to the variability of prices in goods. The first dimension is country-level characteristics. We come up with three factors, country-level GDP, GDP per capita, and the great-circle distance from Beijing to the destination country. The reasoning behind these three factors is that we suspect that sellers may have the incentive to charge a higher premium on their goods to the countries with a relatively high standard of living, as reflected in a high country-level GDP and GDP per capita. Therefore, the price range of the products sold into these countries may be greater than those sold into the countries with a low standard of living. Similarly, we reason that the farther away the country is from China, the greater the price range can be because the sellers are bearing more risks in terms of damaged or lost shipments. The second dimension we seek to explain price dispersion is based on the good characteristics. Based on the paper published by *James E Rauch (1999)*, we group the products into 3 different categories: those sold on organized exchanges, those which are reference priced, and those which are highly differentiated commodities. The products sold on an organized exchange are treated as the least differentiated products and the differentiated commodities are treated as the most differentiated products. The method that Rauch has used to determine the classification of the products is through looking at whether the product possesses a reference price. A possession of a reference price distinguishes homogeneous from differentiated products. In addition, homogeneous products can be further divided into those whose reference prices are quoted on organized exchanges and those whose reference prices are quoted only in trade publications. Our reasoning for examining this factor is that the most differentiated products, such as designer clothes or laptops, are most likely to have higher price dispersion when sold in different countries, whereas the least differentiated products, such as cardboard boxes, are expected to have less price dispersion. Another good-specific characteristic that we seek to explain is capital or labor intensity of the product. To figure out which factor has a bigger influence on the price variation of the product, we calculate the capital/labor intensity ratio and explore the relationship between the ratio and price dispersion. In the end, we also try to figure out the variability of prices across five continents and whether price variability increases over time.

A large part of our research builds upon the research by *Crucini, Telmer, Zachariadis (2005)* that studies European real exchange rate based on the retail data on European countries. The paper studies good-by-good deviations from the Law of One Price and uses local-currency retail prices for a broad set of goods and services in all EU countries over five-year intervals between 1975 and 1990. The authors determined that after controlling for differences in income and value-added tax rates, there are roughly as many overpriced goods as there are underpriced goods between any two EU countries. We use a similar approach when trying to explore price dispersion in international prices of traded goods from China; however, unlike *Crucini, Telmer, Zachariadis (2005)*, we use a different method in calculating price dispersion (using the average price as the numeraire) and focus on explaining price dispersion based on country-specific and good-specific characteristics.

While we are among the first papers in the literature in trying to explain international price dispersion through trade data, the data we use are also used in the paper by *Manova and Zhang (2012)*, which has a different purpose of establishing stylized facts about firm's export prices using this data source. Based on the same data source, Manova and Zhang (2012) conclude that more successful exporters use higher quality inputs to produce higher quality goods, and firms vary the quality of their products across destinations by using inputs of different quality levels.

The remainder of this paper is organized as follows. We will describe the structure and scope of our data source in Section II. Section III will present our basic methodology for calculating International Price Dispersion and the Law-of-One-Price Deviations. Section IV presents the regression results of the paper and the interpretations and possible explanation that come with the results. The last section is the conclusion.

II. The Data

The original source of our data comes from Chinese firms that participated in international trade over the 2000-2006 period. These data have been collected and made available by the Chinese Customs Office. Specifically, the data report the f.o.b value, quantity, and price from firm-level exports across products and destination countries. The dataset includes as many as 213 destination countries and a total of 5144 different products with over 4 million observations and each observation has 26 variables such as product code, firm id, annual quantity exported, annual deflated value, date, and deflated price.

Table 1 presents basic information about the scope of the data. For example, the total number of observations in 2000 is 247,804, which includes 195 countries and 4193 different products. Since different firms can sell the same product to the same destination, we average the prices of all firms (weighted by annual quantity) that traded a specific good to a specific destination and call it the price of a product-destination pair. As a result, we calculated a total of 66,585 distinct product-destination pairs prices in year 2000. According to the number of countries and products mentioned above, we can calculate the possible number of combinations by multiplying the two numbers together. Due to the large number of countries China traded to each year and the number of different product China exports, there are a great number of missing observations in terms of product-destination pairs. However, as the total observations increase by large amount each year, and the number of countries and products observed increase by a much smaller amount, we can see that the proportion of missing product-destination pairs decreases over the 7-year period.

To help our readers understand the dataset more clearly, Table 2 presents the average prices of 10 distinct products in combination with ten most common-traded partners over the 7-year period. For instance, for the product code 950699, which represents equipment for general physical exercise, gymnastics, or athletics, the average price of the 165 observations (not shown in the table) that was traded with Korea equals 4.38 dollars after quantities of each observations are weighted over the 7 years. In addition, as we can infer from the table, there is not a single observation of meat and edible offal traded to Australia or woven fabric of cotton traded to Canada over the period. As is evident from this table, the data is quite comprehensive as it includes many categories such as food, clothing, appliances, metal, chemicals and so on.

Because our paper aims to relate international price dispersion to either country-specific characteristics or good-specific characteristics, we have also supplemented our data with output-side real GDP for each country. Based on the population which is also in the data, we can calculate GDP per capita, which is another country-specific independent variable for our research. The third variable distance provided in the data set is the great-circle distance from the most populated city of the export country to Beijing. For the good-specific characteristics, we have used two different versions of classification systems that classifies the differentiability of the products based on the paper by *James E Rauch (1999).* In particular, because ambiguities arose that were sometimes sufficiently important to affect the classification, both “conservative” and “liberal” classifications were made, was the former minimizing the number of commodities that are classified as organized exchange or reference priced and the latter maximizing those numbers. For example, the product “flexible tubing of iron or steel” may be included in the organized exchange category by the “conservative” standard, but in the reference priced category by the “liberal” standard.

III. Methodology

i. Law-of-One-Price Deviation

In order to calculate the Law-of-One-Price Deviation, we define $Z_{i,j,k}$ to be

$$Z_{i,j} = P_{i,j} / \overline{P}_i ,$$

Where i stands for product, j stands for the destination, and $\overline{P}_i$ is the average price of the product for all firms weighted based on the quantity traded to each country. Therefore, we define log $Z_{i,j,k}$ as the Law-of-One-Price deviation for good i listed by the firm k and sold to the destination j.

ii. International Price Dispersion

In trying to understand the factors that influence international price dispersion, the basic building block of our measure of price dispersion is the percentage deviation of the price of good i traded to country j from the average price of the same good across countries:

$$y_{i,j} = (P_{i,j} / \overline{P}_i) - 1,$$

where $\overline{P}_i$ is calculated using the same methodology as in the Law-of-One-Price deviation calculation. To make the results more accurate, we measure the international price dispersion as

$$k_i = \text{mad}(y_{i,j}),$$

where mad() is the mean absolute deviation and is less sensitive to large outliers than the standard deviation. k_i is therefore the dependent variable in our research that we seek to explain with country-specific and good-specific characteristics.

The country-specific characteristics can be run in the simple regression model in a straightforward way. We use a linear regression model to understand the variation in prices across countries. Specifically, we consider

$$k_i = \alpha + x_i \beta + u_i, \ i = 1, \ldots N,$$

where k_i is price dispersion of good i across countries, x_i is a matrix of explanatory variables, and u_i is an error term assumed to be *i.i.d.*

For the good-specific characteristics, the classification of degree of differentiation is a categorical variable of rank 3, meaning that each good can fall under one of the three categories. When we are trying to explain price dispersion by good-specific characteristics, we will create 2 dummy variables β1 and β2, with the highly-differentiated commodities as the reference group. Therefore, according to the equation

$$k_i = \alpha + x_1 \beta_1 + x_2 \beta_2 + u_i$$

When both x_1 and x_1 equals to 0, the equation represents the inflence of the reference group, which is the highly-differentiated commodities group. When x_1 equals 1 and x_2 equals 0, the coefficient of the reference-priced group can be calculated as the addition of α and β_2; The coefficient of the organized exchange group can be calculated when x_1 equals 0 and x_2 equals 1.

We use a similar method when we are trying to explore the varaibility of prices across regions, continents, and time. For instance, when we are trying to explore how the price variability changes over a 7-year-period from 2000 to 2006, we create 6 dummy variables x_1, x_2, ..., x_6, and the equation becomes

$$k_i = \alpha + x_1\beta_1 + x_1\beta_2 + ... + x_6\beta_6 + u_i ,$$

where α represents the influence of the reference year 2000, and the influence of the other years can be calculated by adding the coeffcient β to the reference year coefficient α.

IV. Results

Figure 1 summarizes the empirical distribution of Law-of-One-Price deviations for each of China's most common trade partner countries. Each chart reports the kernel distribution of log $Z_{i,j}$, which is the Log-of-One-Price Deviation for different products exported to that country over the 7-year period. Figure 1 displays two major features. First, the distributions are all centered around 0, which suggests that for each country, there are roughly as many overpriced exports as there are underpriced exports. Secondly, some of the distributions, Brazil and Canada for example, have a slightly longer tail on the right-hand side of the distribution, which suggests that the extreme prices are more likely to appear above the average price of exports across countries than to appear below the average price.

Table 3 considers the effect of product differentiation on the international price dispersion. As we mentioned earlier, we divide the products into three categories: those sold on organized exchanges, those which are reference priced, and those which are highly differentiated commodities. We used the highly-differentiated commodities as the reference group, and came up with the coefficient for all three groups under two different standards. Under the conservative standard, the coefficients for highly differentiated commodities, those that are referenced priced, and those sold on organized exchange are 35.61, 7.79, and 1.74 respectively, which suggests that the more differentiated products will tend to have a larger price dispersion. This result is very much in line with our expectation. The R-square of this result is 0.03%, meaning that about 0.03 percent of the priced variation can be explained by product differentiation. Although the R-squared is not very high, the F test provided by Stata for each individual group rejects the null hypothesis that the coefficients on the independent variables are equal to 0, which shows that the result is statistically significant.

Similarly, the second part where the classification system is based on a more liberal standard, the coefficients for highly differentiated commodities, those that are referenced priced, and those sold on organized exchange are 35.49, 14.32, and 1.80 respectively. Again, these results are very similar to the results obtained by using the conservative standard and have confirmed our expectation about how product differentiation will affect the international price dispersion. The R-square equals to 0.02%, and the F test again shows that the result is statistically significant.

Table 4 demonstrates the effect of country-level GDP, GDP per capita, and distance from the destination have on price dispersion. Based on the regression results, we observe the striking fact that country-level GDP has a positive influence on the price dispersion whereas GDP per capita has a negative influence; however, the coefficient of country-level GDP (6.71e-13) seems to be too small compared with the coefficient related to GDP per capita (-0.00031). Since both country-level GDP and GDP per capita are denoted in dollars, we think that the influence of country-level GDP has on price dispersion is negligible compared that of GDP per capita. Therefore, the negative coefficient of GDP per capita represents that for each dollar increase on GDP per capita, price dispersion will decrease by 0.00031. In trying to explain this result, we think that GDP per capita, compared with country-level GDP, is a better reflection of standard of living for a particular country. Since countries with relatively low GDP per capita are usually the countries with the largest income inequality (especially countries in the African continent) that contain very few extremely wealthy individuals and mostly poor citizens, the sellers are more likely to charge extra premium on the price of the high-end products or extract the profits to a minimum for some of its low-end products. Therefore, the countries with relatively low GDP per capita might have a large price variation for the goods they imported. In the future, we can further explore this hypothesis by extending our research by looking at whether there is a close connection between GDP per capita and Gini coefficients. In order to explain the negative coefficient associated with the distance, we think that the countries that are relatively close to China are more likely to have exports with a wider product scope, which may in turn make the price variance bigger. For example, tomatoes and other vegetables may be sold into other Asian countries from China, whereas the US decides to import tomatoes mainly from Mexico since it is much cheaper and more convenient to do so. The R-square for these results is low, suggesting that a lot of other variables might be responsible for price dispersion; however, we can conclude that the result is statistically significant based on the F-test that rejects the null hypothesis that the coefficients on the independent variables are equal to 0.

Table 5 shows the regression results regarding the capital/labor intensity of the product. Since the coefficient 0.00581 is positive, it can be interpreted that the higher the capital/labor ratio is, the more likely the product will have a larger price range when sold as an export. This further suggests that products that are more capital-dependent are more likely to have a bigger price dispersion when sold in the international markets. In trying to interpret the result, we think that the positive coefficient might be explained by the fact that the cost of capital is more susceptible to factors outside economic conditions. For example, if a food product has corn in its ingredients, and the price of corn fluctuates, as a result, as the cost of capital are more likely to change dramatically in the market than the cost of labor, the price of more capital-dependent products may vary in a greater extent than the price of more labor-dependent products. Alternatively, we might also interpret the results in the way that capital-intense production process tends to produce more complex products, so there may be more scope for price variation in these industries.

Table 6 shows the regression results for different continents. After comparing the results, we can see that the products sold in Asia have the most price variation while those sold in Europe have the least, with Africa, Americas, and Oceania in the middle. Strikingly, the coefficient of Asia (53.80) more than double that of the Europe (21.30) and of Oceania (23.90), and we suspect that this might be due to the fact that export products with a greater degree of differentiation are more likely to be sold to the nearby Asian countries than to the other continents. Thus, those products might be sold in many varieties and quantities, which can make a big difference in creating a large degree of price variation.

Table 7 presents the regression results from regressing the price dispersion measure on a series of annual dummy variables. As we can observe from the figure, the coefficient steadily increases over the 2000-2006 period, with the annual increases being 3.24, 2.49, 0.15, 3.97, 1.51, and 3.75 from 2000 to 2006. Although the year 2002 to 2003 represents a year with negligible growth in price dispersion, the figure demonstrates a relatively consistent increase in price dispersion for all the products overall. Consequently, the coefficient in year 2006 (44.73) represents almost a 50% increase from the coefficient for the year 2000 (29.63). We think this increase in price dispersion is due to the dramatic increase on the number of price observations from year 2000 (247,804) to year 2006 (1,042,284). Because the number of products (4193 to 4643) and the number of countries (195 to 209) did not increase significantly after 6 years, the number of observations in 2006, which is more than four times the number in 2000, represent a major factor in the much wider price range for different products.

V. Conclusion

This paper explores the nature of international price dispersion among Chinese exports. By first looking at the distribution of Law-of-One-Price distributions, we conclude that there are roughly as many overpriced exports as there are underpriced exports and that the extreme prices are more likely to appear above the average price of exports. Then we aim to figure out how various factors in two dimensions can affect the international price dispersion. From the first dimension—good-specific characteristics, we observe that the more differentiated products will tend to have a higher price dispersion and that products that are more capital-dependent than labor-dependent tend to have a higher price dispersion when traded to different countries. For country-level characteristics, we observe that the standard of living in individual countries, represented by GDP per capita, is inversely related to price dispersion for each country, whereas the great-circle distance to the export destination has a negative influence on the magnitude of price dispersion. In the end, we look at how price dispersion varies across different continents and time periods. We find that the products sold in Asia have the most price variation while those sold in Europe have the least. Overall, price dispersion for all exports increased steadily throughout the 2000-2006 period, and the magnitude of price dispersion increased by 50% at the end of year 2006 from the level of year 2000.

Table I: Scope of the Price Data

Year	2000	2001	2002	2003	2004	2005	2006
Total observations	247,804	300,876	391,184	509,187	768,336	902,899	1,042,884
Number of countries	195	196	204	205	209	210	209
Number of products	4193	4237	4394	4490	4590	4584	4643
Numeber of Observed Product-destination Pairs	66,585	76,966	91,869	107,649	133,800	148,334	162,774
Number of Possible Product-destination Pairs	817635	830452	896376	920450	959310	962640	970387
Proportion missing	91.9%	90.7%	89.8%	88.3%	86.1%	84.6%	83.2%

Notes:
1. Total observations stand for the total number of firm-level price data for each year.
2. Number of Possible Product-destination Pairs= number of countries*number of products.
3. Proportion missing is calculated by 1-Number of Observed Product-destination Pairs/ Number of Possible Product-destination Pairs

Table II: Average export prices across different destinations over 7 years

Code	Good description	AUS	BRA	CAN	CHE	DEU	HKG	IDN	JPN	KOR	USA
020714	Meat and edible offal		1.45	1.45	1.47	1.29	1.17	1.31	1.42	1.45	1.43
190220	Stuffed pasta	2.21	1.30	2.24	2.40	2.11	2.29	1.45	2.05	1.71	2.07
282110	Iron oxides and hydroxides	0.46	0.46	0.48	0.43	0.45	0.57	0.48	0.47	0.50	0.48
392690	Plastics used in machines or instruments	1.89	2.66	1.85	2.37	2.06	2.23	2.36	2.09	2.19	1.98
481920	Folding cartoons, boxes, and cases	1.37	2.12	1.44	1.57	1.48	1.26	0.90	1.70	2.00	1.33
520812	Woven fabric of cotton	0.62		0.48	0.49	0.54	0.55	0.53	0.50	0.52	0.50
620433	Women's or girl's suits, jackets, blazers, dresses, skirts of synthetic fibers	6.10	10.16	8.83	8.68	8.67	4.35	12.14	10.46	8.28	8.11
732690	Iron or steel	1.34	2.05	1.47	1.98	1.56	1.50	1.12	1.39	1.48	1.41
851671	Drip coffee maker	8.42	9.70	8.91	8.08	7.83	8.30	7.78	8.82	7.02	9.00
950699	Equipment for general physical exercise, gymnastics, athletics	4.31	3.95	2.86	4.19	4.44	3.72	4.26	2.60	4.38	2.96

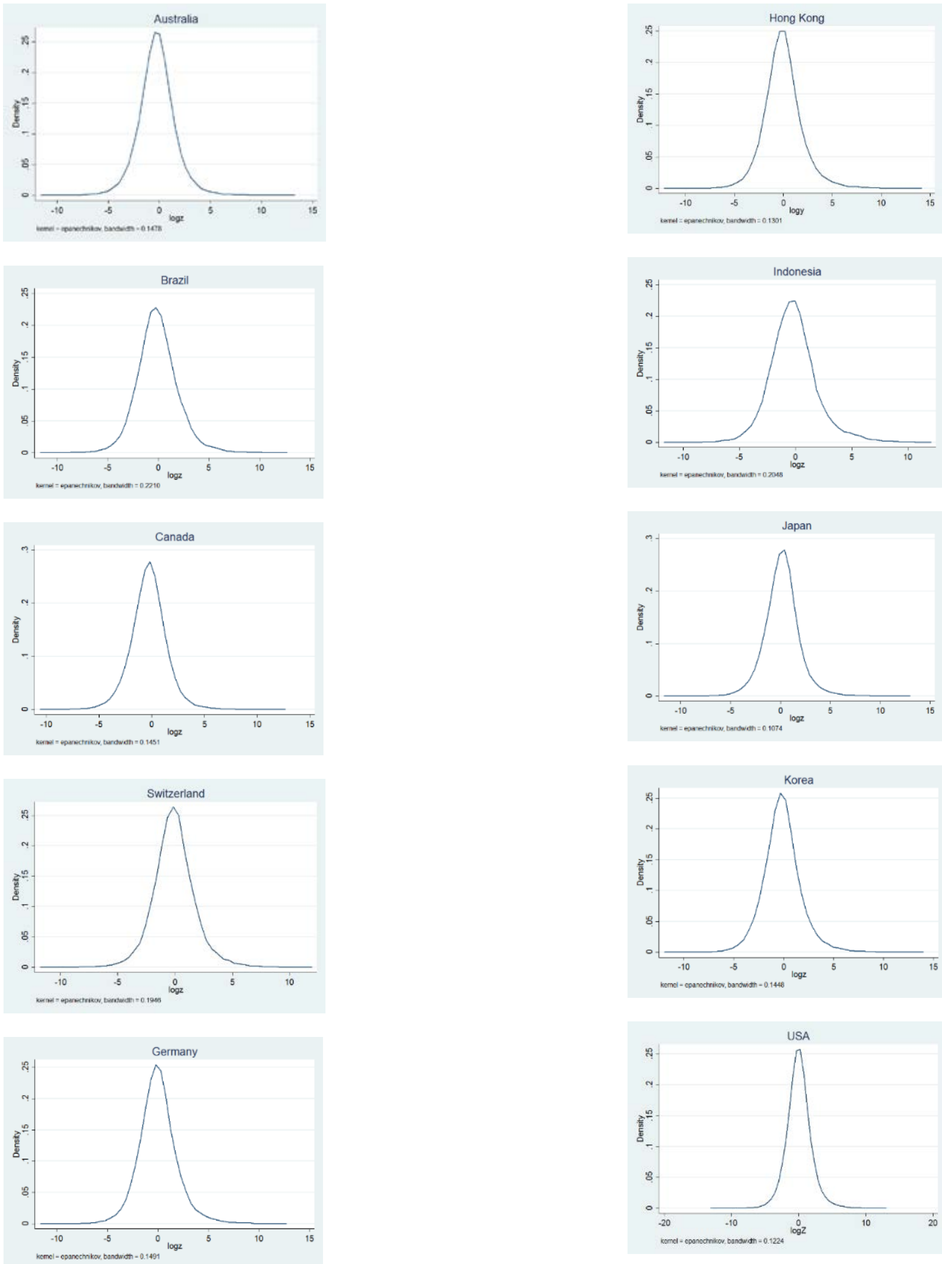
Notes:

1.The country from left to right stands for Australia, Brazil, Canada, Switzerland, Germany, Hong Kong, Indonesia, Japan, Korea, and USA.

2. The prices shown are the average export prices for price-destination pairs calculated based on all the price observations at firm-level.

3. Since these countries are the most common-traded partners of China, only one entry in the table does not have a value because of missing data. If we look at the full set of data, price entries for many product-destination pairs are missing due to lack of data.

Figure 1： Law-of-One-Price Distributions the most Common-traded Country



Notes:

1. The graphs reflect the kernel density estimates of Law-of-One-Price deviations for the most common trade partner countries: Australia, Brazil, Canada, Switzerland, Germany, Hong Kong, Indonesia, Japan, Korea, and USA listed from top to down, left to right.

Table 3: Regression Results across Degrees of Product Differentiation		
Variable	**Coefficient Conservative Standard**	**Coefficient Liberal Standard**
Highly Differentiated Commodities	35.61***	35.49***
	(0.27)	(0.27)
Reference Priced	7.79***	14.32***
	(0.84)	(0.78)
Organized Exchange	1.74	1.80
	(3.10)	(2.40)
R-square	0.0003	0.0002

Notes: ***p < 0.01, **p < 0.05, *p < 0.1. All the products were divided into three categorical variables according to two different sets of standards.

Table 4: Regression Results Using Country-specific Characteristics	
Variable	**Coefficient**
Country-level GDP	6.71e-13***
	(1.02e-13)
GPD per capita	-0.00031***
	(0.00001)
Distance	-0.00303***
	(0.00007)
R-square	0.0004

Notes: ***p < 0.01, **p < 0.05, *p < 0.1. Country-level GDP and GDP per capita are both denoted in dollars measured in 2005. Distance stands for the great-circle distance between the capital cities of two countries, and is denoted in kilometers.

Table 5: Regression Results across Capital/labor Intensity	
Variable	**Coefficient**
Capital/Labor Ratio	0.00604***
	(0.00069)
R-square	0.0000

Notes: ***p < 0.01, **p < 0.05, *p < 0.1. Capital/labor ratio is calculated by dividing the total capital cost of producing the product by the total labor cost on a firm-level basis.

Table 6: Regression Results across Different Continents	
Variable	**Coefficient**
Americas	29.60**
	(1.77)
Asia	53.80***
	(1.67)
Africa	34.68***
	(1.61)
Europe	21.30***
	(1.74)
Oceania	23.90***
	(2.24)
R-square	0.0005

Notes: ***p < 0.01, **p < 0.05, *p < 0.1.

Table 7: Regression Results over Years	
Variable	**Coefficient**
2000	29.63***
	(1.28)
2001	32.85*
	(1.73)
2002	35.35***
	(1.63)
2003	35.49***
	(1.56)
2004	39.46***
	(1.47)
2005	40.97***
	(1.44)
2006	44.73***
	(1.42)
R-square	0.0005

Notes: ***p < 0.01, **p < 0.05, *p < 0.1.